Vers un modèle de canal réaliste pour l'analyse de la sécurité du processus d'authentification par code matriciel 2D

Cléo BARAS, François CAYRE

GIPSA-Lab, département Images et Signal
11 rue des Mathématiques, Grenoble Campus BP46, F-38402 Saint Martin d'Hères Cédex, France
cleo.baras@gipsa-lab.grenoble-inp.fr, francois.cayre@gipsa-lab.grenoble-inp.fr

Résumé – Le problème de l'authentification de produits manufacturés par des codes matriciels 2D (CM-2D) encodant des identifiants binaires est analysé du point de vue d'un contre-facteur cherchant à estimer puis à reproduire le code matriciel à partir de N_c données authentiques, imprimées sur les emballages des produits puis scannées. Ce travail propose un modèle du canal d'impression et de numérisation basé sur des données terrains le plus réaliste possible puis un nouvel estimateur de CM-2D. Trois métriques d'estimation sont testées, la plus adéquate étant une moyenne pondérée calculée sur la zone d'impression de chaque 0 ou 1. Les effets du canal sur cette métrique suivent deux lois log-normales asymétriques, permettant de formaliser la probabilité d'erreur d'estimation et de sélectionner un seuil de décision (entre 0 et 1) qui la rende minimale en fonction de N_c. Des expérimentations sur des données réelles montrent que le contre-facteur peut déjouer le système d'authentification en collectant seulement une dizaine de données authentiques (contre une centaine dans des travaux antérieurs).

Abstract – This paper investigates the authentication problem of real-world goods on which 2D Bar-Codes (BC) encoding binary identifiers were printed. The problem is considered from the opponent's point of view, assuming that he has access to N_c copies of a genuine 2D-BC each distorted by the printing and scanning processes (namely the channel). A channel model is proposed to be the more realist regarding real acquired data. A new estimator of the BC letting the opponent print a fake BC is then deduced. It depends on an estimation metric and a threshold making the decision between estimated bits (0 or 1). Three estimation metric are investigated. It is shown that the channel effects on the considered metrics can be modeled using 2 asymetrical log-normal distributions, letting us choose the most suitable metric/threshold couple, that minimises the estimation error probability. Experimental results are finally given on real data. It is shown that the opponent can produce a fake that successfully fools the authentication system with few dozen of copies (compared to few hundred in previous works).

1 Introduction

L'une des technologies envisagée pour l'authentification de produits manufacturés (tels les médicaments ou les consommables) est un code matriciel sécurisé (CM-2D) [1] imprimé de manière visible sur tous les emballages des unités d'un lot de fabrication (ou copies) d'un produit. Il s'agit d'une image en noir et blanc, qui à la manière d'un tatouage encode un identifiant binaire sur la base de clés secrètes, et est conçue pour se dégrader naturellement à l'impression donc être incopiable [2]. Une fois scanné, le CM-2D est soumis à un détecteur qui décode l'identifiant binaire et décide si le produit provient de la chaîne de fabrication officielle ou a été contre-fait.

Les performances d'un tel système sont étroitement liées à la sécurité du processus d'authentification, *i.e* la facilité qu'aurait un attaquant à fabriquer puis imprimer un CM-2D sur un produit contre-fait qui serait déclaré authentique par le détecteur. Elle est mise à mal dès que le contre-facteur a accès à N_c copies d'un même CM-2D obtenues dans un même lot de fabrication. En effet, elles peuvent lui servir à estimer le CM-2D, sans même avoir besoin d'en décoder l'identifiant binaire. Dans [3], nous avons analysé ce scénario d'attaque vu du contre-facteur et proposé un estimateur de CM-2D. Les résultats obtenus montraient d'importants écarts entre ses performances théorique et pratique, principalement dus à un modèle du canal d'impression et de numérisation des CM-2D très éloigné de la réalité. Les modèles de la littérature se concentrent : 1) sur la dispersion d'un pixel imprimé sur le papier et sont finalement trop fins pour notre problème ; 2) sur des déformations de tapis [4] (en tatouage) au détriment de la nature du bruit d'impression et de numérisation et ne s'appliquent donc pas à notre contexte.

Dans cette contribution [1], nous proposons donc un modèle de canal plus réaliste choisi en analysant les statistiques de trois métriques d'estimation. Un nouvel estimateur de CD-2D est ensuite proposé, paramétrable par un seuil de décision ; ses performances théoriques seront dérivées pour établir le choix optimal de ce seuil. La sécurité du procédé d'authentification est finalement réévaluée aux travers de simulations mettant en évidence le gain apporté par cet estimateur au contre-facteur.

2 Le système d'authentification

Le système d'authentification dans sa branche officielle et dans le circuit de contre-façon est présenté figure 1.

1. Travaux subventionnés par le projet ANR Estampille.

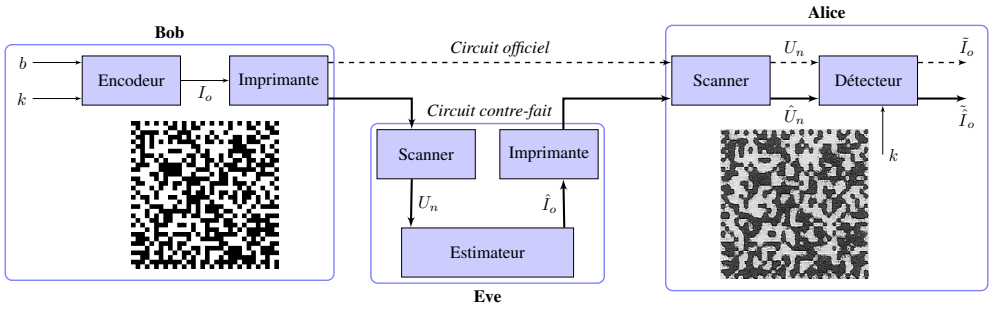

FIGURE 1 – Le système d'authentification. A gauche : un CM-2D. A droite : sa copie zoomée, imprimée sur 16 mm^2 puis scannée avec du matériel grand public (imprimante *HP LaserJet 4250 PS* à 1200 dpi, scanner *HP Photosmart 8200 Series* à 2400 dpi).

2.1 Principes et notations

Un CM-2D (cf. figure 1 à gauche) encode un identifiant binaire en utilisant des codes correcteurs d'erreur (CCE) apportant une certaine robustesse et des embrouillages basés sur des clés secrètes. Il contient également des données de synchronisation qui ne seront pas prises en considération dans cette étude (la synchronisation étant supposée faite). La séquence binaire obtenue est ensuite modulée en amplitude et dispatchée sur une matrice I_o de $M \times M$ éléments noirs (bit 0) ou blancs (bit 1 d'amplitude Δ avec typiquement $\Delta = 255$).

Dans la chaîne de fabrication officielle (cf. figure 1 au centre), le CM-2D est imprimé en petite taille (de l'ordre de 16 mm^2) sur l'emballage des produits. La copie du n-ième produit est scannée puis soumise au détecteur pour décider si elle est authentique ($\mathcal{A}$) ou contre-faite ($\mathcal{C}$). Cette décision se base sur la matrice $U_n = \mathcal{P}_n(I_o)$, n-ième copie du CM-2D ayant subit les dégradations du canal d'impression et de numérisation $\mathcal{P}$. Une interprétation visuelle de ces dégradations (cf. figure 1 à droite) montre : 1) un bruit lié (notamment) à la dispersion de l'encre sur le papier ou aux conditions de prise de vue et 2) un rééchantillonnage du scanner. Chaque élément de U_n prend désormais sa valeur dans l'échelle de niveaux de gris $[0, \Delta]$ et s'étale sur $S \times S$ pixels (avec S non entier, dépendant de la taille du CM-2D, de la résolution d'impression et de numérisation).

Le contre-facteur collecte N_c copies d'un même CM-2D, ayant chacune subie des dégradations différentes liées au canal. Il procède à une estimation $\hat{I}_o$ du CM-2D sur la base des acquisitions U_n qu'il réimprime sur sa contre-façon. A supposer que le contre-facteur ait accès au même matériel d'impression que la chaîne officielle (pire scénario pour le fabricant officiel), le CM-2D $\hat{I}_o$ réimprimé subira lui aussi les dégradations du canal avant d'être soumis au détecteur sous sa forme scannée $\hat{U}_n = \mathcal{P}_n(\hat{I}_o)$ pour décider de son authenticité.

2.2 Performances du système

Les performances d'un tel système sont (entre autre) déterminées par la probabilité de décider qu'une copie est authentique alors qu'elle est contre-faite. Puisqu'elle est fonction de paramètres (dont les clés secrètes) inconnus du contre-facteur, la qualité du CM-2D contre-fait $\hat{I}_o$ ne peut être évaluée que par la probabilité d'erreur $p_{e|\mathcal{C}}$ commise par le détecteur dans le décodage de $\hat{I}_o$ après réimpression. Cette probabilité $p_{e|\mathcal{C}}$ dépend : 1) de la probabilité $p_e(N_c)$ d'erreur dans l'estimation des éléments binaires du CM-2D original I_o à partir de N_c copies collectées, et 2) de la probabilité d'erreur nominale du détecteur p_d inhérente à sa mise en oeuvre. Elle s'écrit [3] :

$$p_{e|\mathcal{C}}(N_c) = \big(1 - p_e(N_c)\big)p_d + p_e(N_c)(1 - p_d). \qquad (1)$$

En pratique, le détecteur est réglé pour tolérer une marge ϵ assurant sa robustesse (à la qualité d'impression, aux conditions de prise de vue) : concrètement, toute copie décodée avec un Taux d'Erreur Binaire (TEB) inférieur à $p_d(1+\epsilon)$ sera déclarée authentique, ϵ étant bien sûr choisi étant données les limites du pouvoir de correction des CCE et des embrouillages. L'objectif du contre-facteur est alors de concevoir un estimateur tel que $p_{e|\mathcal{C}} < p_d(1 + \epsilon)$ assurant à son CM-2D contre-fait $\hat{I}_o$ de déjouer le système d'authentification.

3 Estimateur du contre-facteur

Dans la lignée de [3] et par souci de simplicité, l'estimateur envisagé pour le contre-facteur procède, pour chaque élément (i, j) des $M \times M$ données binaires du CM-2D, en trois étapes : 1) il exploite l'étalement de l'élément sur $S \times S$ pixels pour extraire, pour chacune des N_c copies collectées, une métrique $V_n(i, j)$ (dans $[0; \Delta]$) ; 2) il combine ces N_c métriques $V_n(i, j)$ sous la forme d'une valeur moyenne $\bar{V}(i, j)$; puis 3) il produit une décision dure $\hat{I}_o(i, j)$ (dans $\{0; \Delta\}$) en comparant $\bar{V}(i, j)$ à un seuil de décision τ marquant la frontière entre les éléments noirs (0) et blancs (Δ). Le choix adéquat de la métrique et de τ dépend du canal d'impression et de numérisation $\mathcal{P}$ qu'il convient donc de modéliser le plus finement possible.

3.1 Modélisation du canal $\mathcal{P}$

Nous nous intéressons ici à une modélisation statistique du canal $\mathcal{P}$ donnant la répartition de trois métriques $V_n(i, j)$ différentes pour chaque élement (i, j) : 1) la valeur *moyenne* de l'élément (i, j) sur ses $S \times S$ pixels, 2) sa valeur *médiane* et 3) une *moyenne pondérée* par une fenêtre de Hanning de taille $S \times S$ donnant plus de poids aux pixels centraux et atténuant les effets de la dispersion de l'encre sur le papier.

Ces métriques sont supposées liées à des réalisations de Variables Aléatoires (VA), qui comme le montreront les résultats expérimentaux, suivent deux lois Log-Normales (LN) dont les paramètres dépendent de l'élément encodé dans (i,j) :

$$\begin{cases} V_n(i,j) & \sim \mathcal{LN}(\mu_0, \sigma_0^2), \text{ si } (i,j) \text{ encode un } 0 \\ 1 - V_n(i,j) & \sim \mathcal{LN}(\mu_1, \sigma_1^2), \text{ si } (i,j) \text{ encode un } 1 \end{cases} \quad (2)$$

où $\mathcal{LN}(\mu, \sigma^2)$ désigne la loi de densité de probabilité $p_V(v) = \frac{1}{\sqrt{2\pi}\sigma v} \exp\left(-\frac{1}{2}\frac{(\ln(v)-\mu)^2}{\sigma^2}\right)$.

3.2 Performances théoriques de l'estimateur

Pour chaque élément (i,j), l'estimateur fournit une décision souple $\bar{V}(i,j) = \frac{1}{N_c}\sum_{n=1}^{N_c} V_n(i,j)$ à partir des N_c copies collectées du même CM-2D. Étant données les lois des $V_n(i,j)$ supposés *i.i.d*, et en utilisant l'approche de Fenton-Wilkinson [5], la métrique $\bar{V}(i,j)$ peut être vue comme la réalisation d'une VA suivant deux lois LN asymétriques telles que :

$$\begin{cases} \bar{V}(i,j) & \sim \mathcal{LN}(\bar{\mu}_0(N_c), \bar{\sigma}_0^2(N_c)), \text{ si } (i,j) \text{ encode un } 0 \\ 1 - \bar{V}(i,j) & \sim \mathcal{LN}(\bar{\mu}_1(N_c), \bar{\sigma}_1^2(N_c)), \text{ si } (i,j) \text{ encode un } 1 \end{cases} \quad (3)$$

avec $\bar{\mu}_i(N_c) = \mu_i + \frac{\sigma_i^2 - \bar{\sigma}_i(N_c)^2}{2}$ et $\bar{\sigma}_i^2(N_c) = \ln\left(1 + \frac{e^{\sigma_i^2}-1}{N_c}\right)$ (pour $i = 0$ ou 1). Par souci de lisibilité, nous omettrons par la suite la mention de la dépendance de ces paramètres à N_c.

La métrique $\bar{V}(i,j)$ est ensuite comparée à un seuil τ pour décider du bit estimé : si $\bar{V}(i,j) < \tau$, alors l'élément (i,j) est estimé être un 0 ; sinon, c'est un 1. La probabilité d'erreur p_e commise par le contre-facteur lorsqu'il estime le CM-2D à partir de N_c copies peut alors être approchée par :

$$p_e(\tau, N_c) \approx \frac{1}{2} - \frac{1}{4}\text{erf}\left(\frac{\ln(1-\tau) - \bar{\mu}_1}{\bar{\sigma}_1\sqrt{2}}\right) - \frac{1}{4}\text{erf}\left(\frac{\ln(\tau) - \bar{\mu}_0}{\bar{\sigma}_0\sqrt{2}}\right) \quad (4)$$

où erf est la fonction d'erreur.

Le seuil de décision peut être choisi de manière optimale pour minimiser $p_e(\tau, N_c)$: $\tau* = \arg\min_\tau p_e(\tau, N_c)$ dépend alors uniquement de N_c et des paramètres du modèle de canal. On peut montrer qu'il converge, lorsque N_c tend vers $+\infty$, vers l'unique solution de $[e^{\mu_0+\sigma_0^2/2}; 1 - e^{\mu_1+\sigma_1^2/2}]$ satisfaisant :

$$\left(\ln(\tau) - \frac{2\mu_0 + \sigma_0^2}{2}\right)^2 = \frac{1 - e^{\sigma_0^2}}{1 - e^{\sigma_1^2}}\left(\ln(1-\tau) - \frac{2\mu_1 + \sigma_1^2}{2}\right)^2, \quad (5)$$

qui elle ne dépend que des paramètres du modèle.

4 Résultats expérimentaux

Des résultats expérimentaux obtenus par impression puis numérisation de CM-2D sont présentés ici pour sélectionner la meilleure métrique, établir la pertinence du modèle de canal et les performances de l'estimateur. Le matériel utilisé est dans chaque cas une imprimante de bureau *HP LaserJet 4250 PS* de résolution 1200 dpi et un scanner grand public *HP Photosmart 8200 Series* de résolution 2400 dpi.

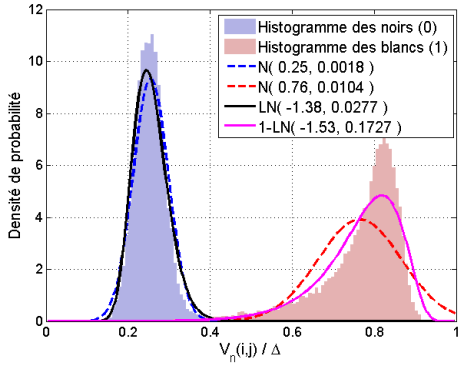

FIGURE 2 – Histogramme normalisé des moyennes pondérées en fonction du type d'élément et comparaison à des distributions normales et log-normales.

TABLE 1 – Adéquation entre histogrammes et lois

Métrique	Moyenne		Médiane		Moy. pondérée	
Elément	0	1	0	1	0	1
D pour loi $\mathcal{N}$	54	41	96	232	14	110
D pour loi $\mathcal{LN}$	22	19	99	175	11	53

4.1 Pertinence du modèle de canal

La figure 2 présente un exemple d'histogrammes pour les éléments 0 et 1 donnant la répartition de la métrique $V_n(i,j)$ dans le cas de la *moyenne pondérée*. Les CM-2D utilisés incluent chacun 32^2 bits et sont imprimés sur 16 mm^2 soit $S^2 \approx 11^2$ Pixels par Element (p/e) encodé. Cette figure les compare à deux lois : l'une Normale (N) comme dans [3] et l'autre LN, dont les paramètres (μ_i et σ_i pour $i = 0$ ou 1) sont estimés à partir des observations et indiqués dans la légende de la figure. Une interprétation visuelle montre que les lois LN sont davantage en adéquation que les gaussiennes avec les histogrammes. Cette observation est confirmée en comparant la distance D entre les histogrammes observés et les distributions envisagées telle que formulée par le test du χ^2 : $D = \sum_{k=1}^{K+1} \frac{(\hat{N}_k - Np_k)^2}{Np_k}$ où $\{p_k\}$ désigne une discrétisation à somme constante de la loi de probabilité p_V en $K+1$ classes, N le nombre de données utilisées pour l'histogramme et $\hat{N}_k$ le nombre de données observées appartenant à la k-ième classe. Les résultats obtenus pour $K = 102$ sont donnés table 1 et montrent une adéquation observations/modèles LN plus importante dans la quasi-totalité des cas que pour le modèle gaussien.

4.2 Performances de l'estimateur

Les performances de l'estimateur sont analysées à l'aide d'un CM-2D encodant 64^2 bits, imprimé 36 fois à différentes tailles,

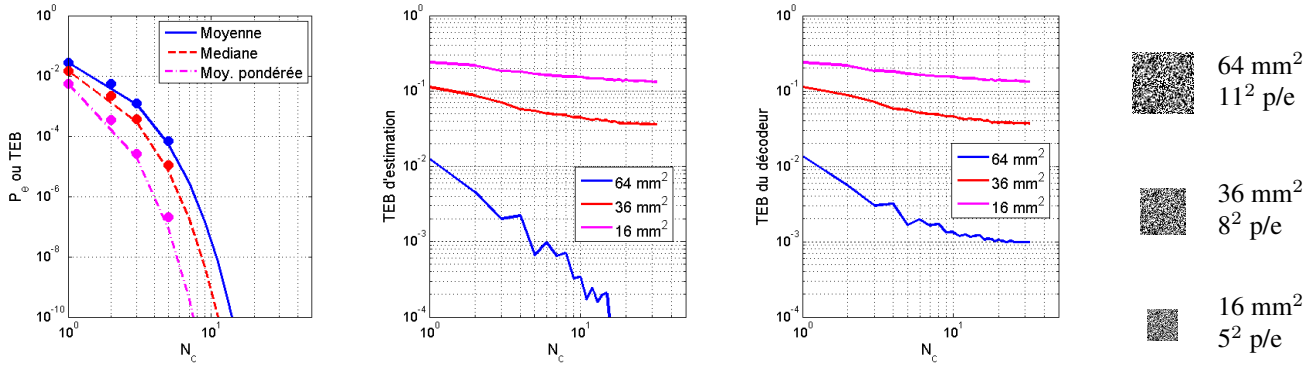

FIGURE 3 – Performances de l'estimateur. De gauche à droite : 1) Probabilité d'erreur (trait) et TEB (rond) théoriques de l'estimateur. 2) TEB de l'estimateur en fonction du nombre N_c de copies collectées d'un même CM-2D pour différentes tailles des copies (en mm^2). 3) Probabilité d'erreur du détecteur lorsqu'il traite le CM-2D contrefait. 4) CM-2D imprimé à différentes tailles.

telles que présentées figure 3 à droite.

La figure 3 de gauche valide l'expression de la probabilité d'erreur $p_e(\tau^*, N_c)$ de l'équation (3) en la comparant aux TEB obtenus en fonction du nombre de copies collectées N_c par une simulation (type Monte-Carlo). Pour chacune des métriques, 10^7 réalisations de $V_n(i,j)$ ont été générées aléatoirement en suivant les modèles LN donnés équation (2). Ces modèles ont pour paramètres les valeurs (μ_i, σ_i) estimées à partir de copies du CM-2D imprimées sur 64 mm^2 soit $S^2 \approx 11^2$ p/e. Le seuil de décision τ^* utilisé est l'optimal minimisant p_e (à N_c fixé). La métrique par *moyenne pondérée* se révèle être la plus efficace et sera la seule analysée par la suite.

La figure 3 présente au centre-gauche les TEB (estimant $p_e(\tau^*, N_c)$) commis par l'estimateur pour construire une estimation $\hat{I}_o$ du CM-2D en fonction du nombre de copies collectées et au centre-droit les TEB du détecteur (estimant $p_{e|\mathcal{C}}$) lorsqu'on lui soumet une copie de $\hat{I}_o$. Ces derniers sont calculés en injectant les TEB expérimentaux de l'estimateur dans l'équation 1 avec une probabilité d'erreur nominale du détecteur de $p_d = 10^{-3}$. On constate que pour des CM-2D de grandes tailles, $\hat{I}_o$ ne diffère pas de I_o de plus de 5 éléments, et ce avec seulement 5 contenus collectés (contre une centaine dans [3]). Ce CM-2D contre-fait met d'autant plus rapidement en défaut le détecteur que sa marge de robustesse ϵ est faible : si $\epsilon = 1$, seuls 5 contenus collectés suffisent à fabriquer un faux détecté comme vrai car de probabilité d'erreur inférieure à $p_d(1+\epsilon) = 2\,10^{-3}$. Lorsque la taille des copies des CM-2D diminue, le TEB d'estimation continue à diminuer avec le nombre de copies utilisées par l'estimateur, mais plus lentement.

Ces résultats expérimentaux confirment les résultats théoriques mais montrent de nouveau des écarts entre théorie et pratique qui proviennent d'erreurs liées : 1) à la synchronisation qui, pour des tailles de CM-2D très petites, devient très sensible ; elle doit notamment faire face aux déformations de tapis induites par les éléments mécaniques de l'imprimante et du scanner, 2) au modèle de la métrique de décision $\bar{V}$ utilisant l'approximation de Fenton-Wilkinson qui perd en précision lorsque N_c augmente.

5 Conclusion

Dans ce papier, un modèle du canal d'impression et de numérisation de codes matriciels 2D utilisés pour l'authentification est proposé : il se base sur une moyenne pondérée calculée sur la zone d'impression de chaque élément binaire du CM-2D et la modélise avec deux loi log-normales asymétriques. Un estimateur du CM-2D est alors proposé, paramétré par un seuil de décision choisi pour minimiser la probabilité d'erreur d'estimation. Cet estimateur montre qu'il est possible, lorsque la taille de traitement des éléments binaires du CM-2D est d'environ 11^2 pixels, de fabriquer un CM-2D contre-fait qui déjoue le système d'authentification en collectant seulement une dizaine de copies authentiques.

Les performances de cet estimateur restent néanmoins limitées par des problèmes de synchronisation, qui deviennent prépondérants à mesure que la taille des CM-2D diminue, et par une erreur d'estimation récurrente liée à des phénomènes de bouchage à l'impression. Ces deux aspects pourront faire l'objet de travaux futurs.

Références

[1] J. Picard, *Digital authentication with copy-detection patterns*. Proceedings of IS&T Optical Security and Counterfeit Deterrence Techniques V, pp. 176-183, 2004.

[2] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld, *Physical one-way functions*. Science, pp.2026-2030, 2002.

[3] C. Baras, F. Cayre, *2D Bar-Codes for Authentication : A Security Approach*. Proc. of EUSIPCO, Romania, 2012.

[4] K. Solanki, U. Madhow, B. Manjunath, S. Chandrasekaran, *Modeling the print-scan process for resilient data hiding*, Proc. SPIE 5681, Security, Steganography, and Watermarking of Multimedia Contents VII, 2005.

[5] L. F. Fenton, *The sum of lognormal probability distributions in scatter transmission systems*, IRE Trans. Commun. Syst., vol. CS-8, pp. 57-67, 1960.